TOWARDS BETTER PERCEPTION OF URBAN INFORMATION: A VISUALIZATION PERSPECTIVE.

by

QIAOMU SHEN

A Thesis Proposal Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in Computer Science and Engineering

October 2019, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

QIAOMU SHEN

TOWARDS BETTER PERCEPTION OF URBAN INFORMATION: A VISUALIZATION PERSPECTIVE.

by

QIAOMU SHEN

This is to certify that I have examined the above Ph.D. thesis

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by

the thesis examination committee have been made.

Prof. Huamin Qu, Thesis Supervisor

Prof. Mounir HAMDI, Head of Department

Department of Computer Science and Engineering

31 October 2019

TABLE OF CONTENTS

Title Page i

Authorization Page ii

Signature Page iii

Table of Contents iv

List of Figures v

List of Tables vi

Abstract vii

Chapter 1 Introduction **1**

1.1 Motivation 1

1.2 Visualization meets urban information 2

1.2.1 Visualization of human movement (People) 2

1.2.2 Visual assist urban environment exploration (Place) 3

1.2.3 Visual interpretation for techniques (Technology) 4

1.3 Contribution 5

1.4 Proposal organization 6

Chapter 2 Conclusion and Future Work **7**

References **8**

LIST OF FIGURES

LIST OF TABLES

TOWARDS BETTER PERCEPTION OF URBAN INFORMATION: A VISUALIZATION PERSPECTIVE.

by

QIAOMU SHEN

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

TBD

CHAPTER 1

INTRODUCTION

1.1 Motivation

The ongoing urbanization process has been one of the most important trend after the World War II (WWII). According to the recent urbanization report from United Nations [?], between 1950 and 2018, the estimated urban population increased more than fourfold, from 0.8 billion to 4.2 billion. The rapid urbanization also result in a series of problems such as the traffic congestion and environmental pollution, which attracts more and more attentions from the research field. Traditionally, it will take tremendous effort for researchers and urban planers to understand and explore the large scale urban dynamics due to the limited resource and existing cases.

Fortunately, the rapid development of information and communication technology(ICT) makes the urban data collection and analysis cheaper and easier, which provides the opportunities for people to study these problems and boost the development of relative discipline such as urban informatics [9] and urban computing [28]. For example, the street view images allow urban planners to conduct urban environment auditing without going to the field; the position data tracking and acquisition techniques embedded in the mobile devices allow researchers to capture the crowd movement pattern of urban residents; industrial emission and the meteorology data collected from the monitoring stations are able to help domain expert in precisely forecasting the air quality in future and provide the control strategy to the government.

Even though the computers have a great advantage in fast computing and large information storage, the fully automatics techniques still have limitations in analyzing the urban data due to the complexity and variety of the real world analysis tasks. It is always need the involvement of human beings who have acute perception and considerable experience in the domain field to make the final decision. Visualization, "the study of transforming data, information, and knowledge into interactive visual representations" [15], bridges the gap between the human beings and the computing techniques.

In this chapter, we first introduce the existing visualization techniques in urban data exploration. Then we discuss the major contributions in this proposal which is followed by the proposal organization.

1.2 Visualization meets urban information

The analysis of model urban always refers to the intersection of three domains: people, place and technology [9], which is covered by the proposed work in this proposal.

- **People.** People refers to the residents, citizens, community groups as well as the relative data such as commuting, social network, etc.
- **Place.** Place refers to the physical environment such as the urban sites, locales and habitats which if formed by nature or human activities.
- **Technology.** Technology refers to a variety of techniques span informatics, computing techniques, wearable devices, etc.

Before introducing the contribution of this proposal, we first briefly introduce the existing techniques from these three aspects:

1.2.1 Visualization of human movement (People)

As a general topic in visualization discipline, how to effectively visualize the human movement has been long studied in recent years. An overview has been proposed to categorized the existing movement visualization techniques into three classes: **pattern extraction**, **direct depiction**, **summarization** [3, 27].

The **pattern extraction** methods leverage effective visualization to discover the mobility pattern from the movement data such as the traffic congestions [24], route assessment [12, 23], commuting patterns [4, 21], co-occurrence [17, 27]. Compare to other two types of techniques, the pattern extract methods aim to solve the specific application problem directly and these techniques are always difficult to extend to other applications.

The **direct depiction** directly plot the movement data according to the data format: such as the line segments for OD trails, and polylines for trajectories [3, 8, 13]. Such techniques preserve all the information of individual movement. However, when the dataset is large, the visual clutter the render workload will be the major limitation which hinders the analysts' ability to explore the data.

The **summarization** techniques aim to provide an overview of movement data. Most of these summarization techniques conduct statistical calculation and aggregation before rendering. The typical visualization can be heatmap [25], density map [14], flow map [11], matrix [26] and node-link [22]. Due to the ability to present the high-level perception and alleviate the data uncertainty [3], the summarization techniques are widely used in movement visualization.

1.2.2 Visual assist urban environment exploration (Place)

The rapid urbanization cause a series of changes of living environment including the pollutant and urban forms. Visualization techniques has been used in these domain to discover the potential problems and support the decision making. Since the different application have significantly different analysis requirement, most of the existing work target at specific tasks.

By combining the data mining and cluster visualization, HydroQual [1] is proposed to support a visual analysis of river quality. To reason the air pollutant in Hong Kong, Qu et al. [18] propose a viusal analytics system consists of several novel design including circular pixel bar chart and a parallel coordinates with S-shape axis. Further more, Deng et al. [6] propose a visual analytics system assisting the air quality analysts in exploring the air pollutant propagation patterns.

Visual analytics also enable the urban planers to design and evaluate the city construction. For example, Ferreira et al. [7] propose Urbane, a 3D multi-resolution framework, which enable the urban planers to conduct the urban development in a data-driven way. Zeng et al. [**?**] proposal a tool to facilitates the urban vitality. Such tool presents both of the physical entities and urban design metrics and allow the expert to quickly discover the city blocks which could be improved.

1.2.3 Visual interpretation for techniques (Technology)

The computing techniques play an important role in the exploration of urban information. Even though more and more techniques are adapted to urban data exploration, most of them are sill used as a black box. For most of the technique developers and users, the interpretability is an important property of the techniques in a variety of real world applications. The interpretation can help the technique developers to improve the technique performance and help the technique users improve their confidence [20].

The technique interpretation can be classified into two categories: model reduction and feature contribution. Model reduction methods usually learn a surrogate model to approximate the original complex model. The surrogate model is usually simple and interpretable, such as linear regression [19] and decision trees [5]. Depending on the ways of approximating the original model's behaviors, there are three main ways to conduct model reduction: decompositional, pedagogical, and eclectic [2].

Feature contribution methods help users understand the relationships between input features and the output prediction. They usually assign each feature an importance score to indicate how it impacts the final prediction. One classical work is Partial Dependence Plot (PDP) [10], which depicts how feature value changes affect predictions. A PDP is usually visualized as a line chart in which the x-axis represents feature values and the y-axis represents prediction possibilities (partial dependence scores). For each feature, its partial dependence scores are usually calculated by iteratively fixing the feature input of all the data points to a certain value and getting the average prediction. One recent work, SHAP [16], also calculates feature attribution, but from a local perspective.

Even though variety of techniques has been proposed to interpret the mechanism of the complex techniques, very few of them target at the application of the urban information exploration which always take large amount, high-dimensioanl and complex data as input.

1.3 Contribution

In this proposal, we introduce several visualization techniques in assisting urban data exploration tasks as well as the understanding of urban computing techniques:

- **Visual Exploration of Human-Scale Urban Forms**. We propose StreetVizor, an interactive visual analytics system that helps planners leverage their domain knowledge in exploring human-scale urban forms based on street view images. Our system presents two-stage visual exploration: 1) an AOI Explorer for the visual comparison of spatial distributions and quantitative measurements in two areas-of-interest at city- and region-scales; 2) and a Street Explorer with a novel parallel coordinate plot for the exploration of the fine-grained details of the urban forms at the street-scale. We integrate visualization techniques with machine learning models to facilitate the detection of street view patterns.
- **An edge bundling technique for visualizing origin-destination trails in urban traffic.** We propose RAEB(route aware edge bundling), a novel edge bundling technique to viusally summary the OD trails in urban traffic data. We identify inconsiderate settings of conventional kernel density estimation edge bundling (KDEEB) when applied to urban traffic data, including non-optimal kernel size and road neglect. The limitations are addressed in RAEB by introduction of a comprehensive pipeline comprising preprocessing, bundling and evaluation processes. A series of new parameters, together with adaptions of existing ones, are employed in the pipeline.
- **Understanding recurrent neural networks on multi-dimensional time-series forecast.** We propose MultiRNNExplorer, a visual analytics system to interpret RNNs on multi-dimensional time-series forecasts. Specifically, to provide an overview to reveal the model mechanism, we propose a technique to estimate the hidden unit response by measuring how different feature selections affect the hidden unit output distribution. We then cluster the hidden units and features based on the response embedding vectors. Finally, we propose a visual analytics system which allows users to visually explore the model behavior from the global and individual levels.

1.4 Proposal organization

The rest of this proposal is organized as follows:

Chapter 2 introduces the design of Streetvizor: specific analysis requirements are described by a collaborating urban planner, and the designs are evaluated and refined against requirements. We illustrate the applicability of our approach with case studies on the real-world datasets of four cities, i.e., Hong Kong, Singapore, Greater London and New York City. Interviews with domain experts demonstrate the effectiveness of our system in facilitating various analytical tasks.

Chapter 3 presents the details about the implementation of RAEB(route aware edge bundling), a new technique to visualize the origin-destination trails in traffic data by edge bundlings techniques. In the evaluation, we conducted experiments with artificial and real-world traffic data to demonstrate the advancements of RAEB in various applications, including the preservation of road network topology and support of multi-scale exploration.

Chapter 4 proposes a MultiRNNExplorer, a visual analytics system to interpret the recurrent neural networks on multi-dimensional time-series forecast. We introduce the system design and demonstrate the effectiveness of the proposed technique with case studies of air pollutant forecast applications.

Chapter 5 concludes the proposal and discusses possible future work.

CHAPTER 2

CONCLUSION AND FUTURE WORK

REFERENCES

[1] Pierre Accorsi, Nathalie Lalande, Mickäel Fabrègue, Agnès Braud, Pascal Poncelet, Arnaud Sallaberry, Sandra Bringay, Maguelonne Teisseire, Flavie Cernesson, and Florence Le Ber. Hydroqual: Visual analysis of river water quality. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 123–132. IEEE, 2014.

[2] Robert Andrews, Joachim Diederich, and Alan B Tickle. Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-based Systems*, 8(6):373–389, 1995.

[3] Gennady Andrienko, Natalia Andrienko, Peter Bak, Daniel Keim, and Stefan Wrobel. *Visual analytics of movement*. 2013.

[4] Roger Beecham, Jo Wood, and Audrey Bowerman. Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems*, 47:5–15, 2014.

[5] Mark Craven and Jude W Shavlik. Extracting tree-structured representations of trained networks. In *Advances in Neural Information Processing Systems*, pages 24–30, 1996.

[6] Zikun Deng, Di Weng, Jiahui Chen, Ren Liu, Zhibin Wang, Jie Bao, Yu Zheng, and Yingcai Wu. Airvis: Visual analytics of air pollution propagation. *IEEE transactions on visualization and computer graphics*, 2019.

[7] Nivan Ferreira, Marcos Lage, Harish Doraiswamy, Huy Vo, Luc Wilson, Heidi Werner, Muchan Park, and Cláudio Silva. Urbane: A 3d framework to support data driven decision making in urban development. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 97–104. IEEE, 2015.

[8] Nivan Ferreira, Jorge Poco, Huy T Vo, Juliana Freire, and Cláudio T Silva. Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, 2013.

[9] Marcus Foth, Jaz Hee-jeong Choi, and Christine Satchell. Urban informatics. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 1–8. ACM, 2011.

[10] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

[11] Diansheng Guo and Xi Zhu. Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2043–2052, 2014.

[12] Xiaoke Huang, Ye Zhao, Chao Ma, Jing Yang, Xinyue Ye, and Chong Zhang. Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE transactions on visualization and computer graphics*, 22(1):160–169, 2015.

[13] Robert Krüger, Dennis Thom, Michael Wörner, Harald Bosch, and Thomas Ertl. Trajectorylenses–a set-based filtering and exploration technique for long-term trajectory data. In *Computer Graphics Forum*, volume 32, pages 451–460. Wiley Online Library, 2013.

[14] Joel Lanir, Peter Bak, and Tsvi Kuflik. Visualizing proximity-based spatiotemporal behavior of museum visitors using tangram diagrams. In *Computer Graphics Forum*, volume 33, pages 261–270. Wiley Online Library, 2014.

[15] Shixia Liu, Weiwei Cui, Yingcai Wu, and Mengchen Liu. A survey on information visualization: recent advances and challenges. *The Visual Computer*, 30(12):1373–1393, 2014.

[16] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, pages 4765–4774, 2017.

[17] Bing Ni, Qiaomu Shen, Jiayi Xu, and Huamin Qu. Spatio-temporal flow maps for visualizing movement and contact patterns. *Visual Informatics*, 1(1):57–64, 2017.

[18] Huamin Qu, Wing-Yi Chan, Anbang Xu, Kai-Lun Chung, Kai-Hon Lau, and Ping Guo. Visual analysis of the air pollution problem in hong kong. *IEEE Transactions on visualization and Computer Graphics*, 13(6):1408–1415, 2007.

[19] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM, 2016.

[20] Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister, and Alexander M Rush. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):667–676, 2018.

[21] Tatiana Von Landesberger, Felix Brodkorb, Philipp Roskosch, Natalia Andrienko, Gennady Andrienko, and Andreas Kerren. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):11–20, 2015.

[22] Tatiana von Landesberger, Felix Brodkorb, Philipp Roskosch, Natalia Andrienko, Gennady Andrienko, and Andreas Kerren. MobilityGraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):11–20, 2016.

[23] Fei Wang, Wei Chen, Feiran Wu, Ye Zhao, Han Hong, Tianyu Gu, Long Wang, Ronghua Liang, and Hujun Bao. A visual reasoning approach for data-driven transport assessment on urban roads. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 103–112. IEEE, 2014.

[24] Zuchao Wang, Min Lu, Xiaoru Yuan, Junping Zhang, and Huub Van De Wetering. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168, 2013.

[25] Leland Wilkinson and Michael Friendly. The history of the cluster heat map. *The American Statistician*, 63(2):179–184, 2009.

[26] Jo Wood, Jason Dykes, and Aidan Slingsby. Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal*, 47(2):117–129, 2010.

[27] Wenchao Wu, Jiayi Xu, Haipeng Zeng, Yixian Zheng, Huamin Qu, Bing Ni, Mingxuan Yuan, and Lionel M Ni. Telcovis: Visual exploration of co-occurrence in urban

human mobility based on telco data. *IEEE transactions on visualization and computer graphics*, 22(1):935–944, 2015.

[28] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38, 2014.